The post-transcriptional trans-acting regulator, TbZFP3, co-ordinates transmission-stage enriched mRNAs in Trypanosoma brucei

Published online 2 December 2011 *Nucleic Acids Research, 2012, Vol. 40, No. 7* ***2869–2883***
doi:10.1093/nar/gkr1106

The post-transcriptional trans-acting regulator, *TbZFP3*, co-ordinates transmission-stage enriched mRNAs in *Trypanosoma brucei*

Pegine B. Walrad, Paul Capewell, Katelyn Fenn and Keith R. Matthews*

Centre for Immunity, Infection and Evolution, Institute of Immunology and Infection Research, School of Biological Sciences, King's Buildings, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK

Received October 12, 2011; Revised November 3, 2011; Accepted November 4, 2011

ABSTRACT

Post-transcriptional gene regulation is essential to eukaryotic development. This is particularly emphasized in trypanosome parasites where genes are co-transcribed in polycistronic arrays but not necessarily co-regulated. The small CCCH protein, *Tb*ZFP3, has been identified as a *trans*-acting post-transcriptional regulator of Procyclin surface antigen expression in *Trypanosoma brucei*. To investigate the wider role of *Tb*ZFP3 in parasite transmission, a global analysis of associating transcripts was carried out. Examination of a subset of the selected transcripts revealed their increased abundance through mRNA stabilization upon *Tb*ZFP3 ectopic overexpression, dependent upon the integrity of the CCCH zinc finger domain. Reporter assays demonstrated that this regulation was mediated through 3′-UTR sequences for two target transcripts. Global developmental expression profiling of the cohort of *Tb*ZFP3-selected transcripts revealed their significant enrichment in transmissible stumpy forms of the parasite. This analysis of the specific mRNAs selected by the *Tb*ZFP3mRNP provides evidence for a developmental regulon with the potential to co-ordinate genes important in parasite transmission.

INTRODUCTION

The core machinery regulating mRNA stability and translation is well conserved throughout eukaryotic evolution, combining with transcriptional control factors to govern the overall expression of a gene (1,2). In addition to the core machinery, a myriad of organism- and tissue-specific *trans*-acting factors co-operate to finesse the genetic regulatory control of development. In recent years, the contribution of *trans*-acting factors that operate to control post-transcriptional, rather than transcriptional, processes have taken increasing prominence in our understanding of gene expression mechanisms. These exhibit regulation at the level of both individual genes and gene networks (3,4). Understanding the complexity of the underlying regulatory signals and machinery nonetheless remains a significant challenge.

One excellent model for the analysis of the post-transcriptional control of gene expression is the kinetoplastid parasites (5,6). These organisms are significant pathogens of the developing world and include *Trypanosoma brucei* (causing Human African Trypanosomiasis; HAT), *Trypanosoma cruzi* (causing South American Chagas' disease) and *Leishmania* spp. that cause a variety of cutaneous and visceral maladies worldwide. Evolutionarily, the kinetoplastid parasites are among the earliest diverged eukaryotic organisms and exhibit a number of characteristics that distinguish them from the Opisthokont model organisms. In particular, their genome is organized into long polycistronic transcription units in which multiple genes are co-transcribed from dispersed unconventional transcriptional start sites (7,8). Despite their co-transcription, however, gene components of these post-transcriptional arrays often display differential expression, such as during the distinct developmental transitions that characterize the progression of kinetoplastid parasites through their complex life cycles (9–12). This differential expression is inevitably controlled at the post-transcriptional level, with regulatory signals being identified predominantly in the 3′ untranslated region (UTR) (13,14) but also present in the 5′-UTR (14,15) and coding region (16) of several experimentally characterized genes.

Perhaps the best characterized models for gene expression control in kinetoplastids are the *procyclin* genes of *T. brucei*. These genes, comprised of *EP1*, *EP2*, *EP3* and *GPEET* isoforms, encode the major surface proteins on the parasite in the midgut of the tsetse fly (17,18),

*To whom correspondence should be addressed. Tel: +441316513639; Fax: +441316513670; Email: keith.matthews@ed.ac.uk

the vector for trypanosomiasis. These proteins differ slightly in their 3′-UTR sequences, which control their differential expression (19–22). Only recently have regulatory *trans*-acting proteins been identified that govern the differential expression of *procyclin* isoforms (22,23). The first of these, *Tb*ZFP3, is one of a family of small CCCH proteins (*Tb*ZFP1, *Tb*ZFP2, *Tb*ZFP3), which are conserved in kinetoplastids with each being implicit in trypanosome differentiation from mammalian bloodstream to tsetse midgut forms (24,25). Specifically, ectopic overexpression of *Tb*ZFP3 elevates the level of EP1 Procyclin protein expression at the expense of GPEET. Moreover, in RNA-immunoprecipitation experiments, *Tb*ZFP3 specifically selects the *EP1 procyclin* mRNA isoform, this being dependent upon both a negative regulatory element (Loop II) in the *EP1* 3′-UTR and the CCCH domain of *Tb*ZFP3, a predicted zinc finger involved in RNA binding in a range of eukaryotic proteins (26–29). Importantly, *Tb*ZFP3 promotes but is not necessary for the translation of the *EP1* transcript, as deletion of the Loop II element suffices to both eliminate the *Tb*ZFP3 interaction as well as grossly upregulate the transcript and protein. This predicts that *Tb*ZFP3 competes with a negative regulator binding to the Loop II element and thereby acts as an anti-repressor to stabilize *EP1* and promote its translation.

These analyses identified *Tb*ZFP3 as the first sequence-specific *trans*-regulator of surface coat regulation in trypanosomes. However, insight into the wider network of regulatory interactions involving this key regulator is lacking. Here, we have carried out a global analysis of the mRNAs that interact with the *Tb*ZFP3mRNP, revealing a role for this regulator in the developmental events associated with parasite transmission from the mammalian bloodstream form to tsetse midgut form. Our findings generate a model whereby a cohort of developmentally regulated genes are co-stabilized in preparation for the signal to differentiate.

MATERIALS AND METHODS

Trypanosomes

SDM-79 medium (30) was used to culture procyclic form *T. brucei*. Transfected cells lines expressing inducible *Tb*ZFP3-TY, *Tb*ZFP3CCAH-TY or *Tb*ZFP3 (NoTag) were described previously (22). Cells were harvested in logarithmic phase growth at $2–6 \times 10^6$ cells/ml. Logarithmic procyclic s427–449 stage cells were induced for ectopic expression using 1 μg/ml tetracycline. RNA and protein samples were harvested simultaneously for all experiments using previously described procedures (31). Stumpy, intermediate and slender form parasites were AnTat1.1 and EATRO *T. brucei*. They were derived using procedures previously described (9).

RNA immunoprecipitation

Anti-*Tb*ZFP3 RNA IPs and western blot analyses were conducted as described (22). Transcripts isolated from four separate, verified, RNA IPs were purified via Qiagen RNAeasy columns (isolating ≥200 nt) and DNAse treated as per the manufacturer's instruction, then pooled for Illumina digital-tag expression analysis.

Illumina digital-tag expression analysis

Transcripts isolated from either anti-*Tb*ZFP3 RNA IPs or whole procyclic cell lysate were reverse transcribed and subject to Illumina digital-tag sequencing by the 'Gene Pool' facility at Edinburgh University (genepool.bio.ed.ac.uk) and by MWG Eurofins. Sequence identities of $\sim 5 \times 10^5$ (RNA IP) and 1×10^6 (total mRNA) quality reads were determined using the *T. brucei* 927 ORF and UTR sequences available from the TriTrypDB database website [tritrypdb.org; (32)].

Transcript stability analysis

Parental and transfected logarithmic procyclic stage cells were induced or uninduced for ectopic '*Tb*ZFP3-TY' or '*Tb*ZFP3-NoTag' expression with tetracycline for 72 h then treated with 5 μg/ml actinomycinD. RNA was harvested at 0 h, 30, 60, 90, 120, 240, 360 and 480 min and prepared via RNAeasy Qiagen columns (isolating ≥200 nt) as per the manufacturer's instructions.

Quantitative RT-PCR

Quantitative RT-PCR was conducted as described previously (22) using the following primers to amplify regions of candidate transcripts: *Tb*SmB (*Tb*927.2.4540): CTTCACAACATCAACCGCAC, CTAACTTCTCCCG AGTTGCG; *Tb*GrpE (*Tb*927.6.2170): CTCTGTTGCTC CAGTCTCCC, TCCAAACCTCTCTCAAGCG; *Tb*R BP23 (*Tb*927.10.11270): ATGGTGTTCTACAGGTCG CC, ACCCGACGTTTCACAAGTTC; *Tb*CYC7 (*Tb*927 .6.5020): TCCCATTTGATGAGGACACATG, GGGAA ACCTGCAACGAATAACCTTCG; *Tb*MCP (*Tb*11.03 .0870): ATACGTGGCTACGGCGTTACG, AGCGGA GCTTAAACCACAGACG; *Tb*KREPA5 (*Tb*927.8.680): CGAATGGAGAGGAGGTTGAG, CACTCCAACGT AGCGACTG; *Tb*SmF (*Tb*09.211.1695): GTGGAAAAG GCAAAACAGGAGACG, GTGCAAGAAATGGAAA GG; *Tb*RPB7 (*Tb*11.01.6090): GAATTCTTCGGCGGT C, TAAACCCACTCATTGCC; *Tb*RPB8 (*Tb*11.02 .5790): ATTGGAAGACACCTTCACGG, GCATGGC ATAATCACAATCG; *Tb*927.5.720: TGCTAATGATC AGTCGCTGG, TTGCGTTGGACACTTCTCAC; *Tb*927.8.1230: TGACGCACTGTTCACTCACAC, CTT GTCGGCGTTATATCGGTC; *Tb*NMD3 (*Tb*927.7.970): AGTTGCAAGAGGTGGTGC, CATTCTCAGGGTTG GC; *Tb*Sm15K (Tb927.6.4340): GTTTCCTTTCCTTCT TGGGC, TTCCTGATGCTACCCCTACG.

Primers for *actin, ep1* and *zfp1* were as described previously (22).

Western and northern blot analyses

Western and northern blots were as described with quantifications derived by G:Box Chemi analyses (Syngene). Probes to the *ORF*s of target transcript candidates and controls were created using Roche DIG RNA Labelling Kit (Sp6/T7) according to the manufacturer's instructions. Primers were as above.

CAT ELISA assay

3′-UTR regulatory regions of candidate target transcripts were PCR-amplified and cloned into the BamHI/XhoI sites of the pHD617 CAT reporter vector from which the Tet-operator sequences had been removed (a gift of P. Macgregor, University of Edinburgh). Primers to amplify and insert the intergenic regions of target candidates were as follows: *Tb*RBP23: GGATCCAAGCTAGA TAAATTAAGTAGTCG, CTCGAGGACCACCCTTT CTCAACAGGCTC, 255 bp 3′ of ORF; *Tb*SmB: GGAT CCGTGGGATCCTTTTACCATC, CTCGAGAATTCC CCTTCACAC, 747 bp 3′ of ORF; *Tb*GrpE: GGATCCC GTAGTTAAATGTGCTTCCG, CTCGAGGGTTCAT TATGGTTCAGC, 1.2 kb 3′ of ORF.

CAT-expression vectors were transfected into procyclic form s427-parasites containing the pHD449 plasmid for tetracycline regulated expression (33), with and without the pHD451 vector (33) for tetracycline inducible expression of 'ZFP3 No Tag', 'ZFP3-TY' and 'ZFP3ccAh-TY' (22). Transfectant cell lines were selected for growth in 1 μg/ml Puromycin and retested for inducibility of ectopic *Tb*ZFP3 expression. Cell lines were induced for ectopic expression in 1 μg/ml Tetracycline for 72 h and protein, RNA and CAT samples were harvested simultaneously. CAT protein expression was examined using the Roche CAT ELISA kit according to the manufacturer's instructions. Trypanosome culture (1 ml) was concentrated by microfuge centrifugation. After removal of the supernatant, the cell pellet was washed three times with cold PBS, before being lysed with 1 ml of Roche lysis buffer for 25 min at room temperature. After removal of cell debris by centrifugation, 500 μl aliquots were snap frozen and stored at −80°C. Samples were measured at three dilutions in duplicate at 1, 3, 5, 7, 10, 12.5, 15 and 20 min within standard parameters, generated using a CAT calibration curve ($r^2 = 0.995$ or above). Results were consistent between all dilutions and time points.

Immunofluorescence microscopy

Procyclic 427 cells with eYFP-labelled Scd6 were washed three times in PBS, then divided equally between SDM-79 media at 27°C, SDM-79 media at 41°C (heat shock) and fresh PBS at 27°C (2 h glucose starvation). Parasites were then fixed in 4% paraformaldehyde for 20 min, washed three times in PBS, quenched in PBT (Triton X-100):glycine (10 mg/ml), washed three times in PBS:BSA and settled onto polylysine-labelled slides. Slides were then incubated with α-*Tb*ZFP3 [1:500] or fresh PBS:BSA 1 h. Slides were washed three times in PBS, incubated in anti-rabbit Alexa-633 [1:5000] (Invitrogen) in PBS:BSA 1 h, washed two times in PBT, stained with 4′,6-diamidino-2-phenylindole, washed two times in PBS and mounted in Mowiol:Phenylene diamine for confocal imaging. Confocal images were captured on a LEICA SP5 confocal microscope and analysed via VOLOCITY (Perkin Elmer).

In situ hybridization

Stumpy form AnTat 1.1 cells were harvested from mice 6-day post-inoculation, examined to confirm morphology, then fixed fresh from blood purification 20 min in 4% paraformaldehyde (pH 7.5), centrifuged at 700 *g* 10 min, washed three times in PBS, quenched in PBT (2% Triton X-100):10 mg/ml glycine and allowed to settle onto polyK-labelled slides 20 min in a humidity chamber. Slides were then washed three times in PBS:10% blocking reagent (Roche), blocked 2 h in hybridization buffer (5 × SSC, 50% formamide, 2% block, 0.02% SDS) and hybridized overnight in either the sense or antisense DIG-labelled oligo probes or hybridization buffer alone overnight in a humidity chamber. The slides were then washed once in 4 × SSC:10% formamide, twice in 4 × SSC, once in 2 × SSC, once in PBS, blocked for 1 h in PBT (0.1% Triton X-100):10% block, incubated 1 h with sheep α-DIG [1:6000] and/or α-*Tb*ZFP3 [1:500], washed four times in PBT (0.1% Triton X-100), twice in PBS, incubated 1 h with $\alpha-\text{sheep}^{\text{Alexa 680}}$ and $\alpha-\text{rabbit}^{\text{Alexa 488}}$ [1:2500], then washed twice in PBT (0.1% Triton X-100), three times in PBS, DAPI stained, washed four times in PBS and mounted in Mowoil:PDA [10:1].

Transcript expression analysis

To highlight stage-regulated concentration distinctions of genes that associate with *Tb*ZFP3, a figure was generated using the EdgeR Bioconductor package for R (34). Comparative transcript concentrations between stumpy and slender form *T. brucei* was plotted with the 'plotSmear' function of EdgeR using $\log^2$ fold change against the $\log^2$ mean count for each mRNA. Additionally, the top 100 transcripts in the *Tb*ZFP3mRNP were highlighted in black.

RESULTS

Selection of transcripts that co-associate with the *Tb*ZFP3mRNP

We used a previously optimized RNA immunoprecipitation (RIP) approach to select mRNAs that associate with the *Tb*ZFP3mRNP (22). As a control, extracts were incubated alternatively in the presence of a peptide matching the epitope recognized by the *Tb*ZFP3-specific antibody (Figure 1A). Analysis of the immunoprecipitated material from four replicate extractions confirmed that there was effective selection of *Tb*ZFP3, this being efficiently blocked in the presence of peptide. Associating RNA was then isolated from the combined extracts derived in the absence of blocking peptide, and the extracted polyA^+ mRNAs subjected to Digital-Tag (Solexa) gene expression analysis on an Illumina platform. The resulting reads were then aligned to *T. brucei* TREU927/4 open reading frames and to a data set of untranslated regions generated by RNAseq analysis of *T. brucei* s427 procyclic form trypanosomes (8,10). This allowed the identification and quantitation of transcript tags from cDNAs either in their coding region or

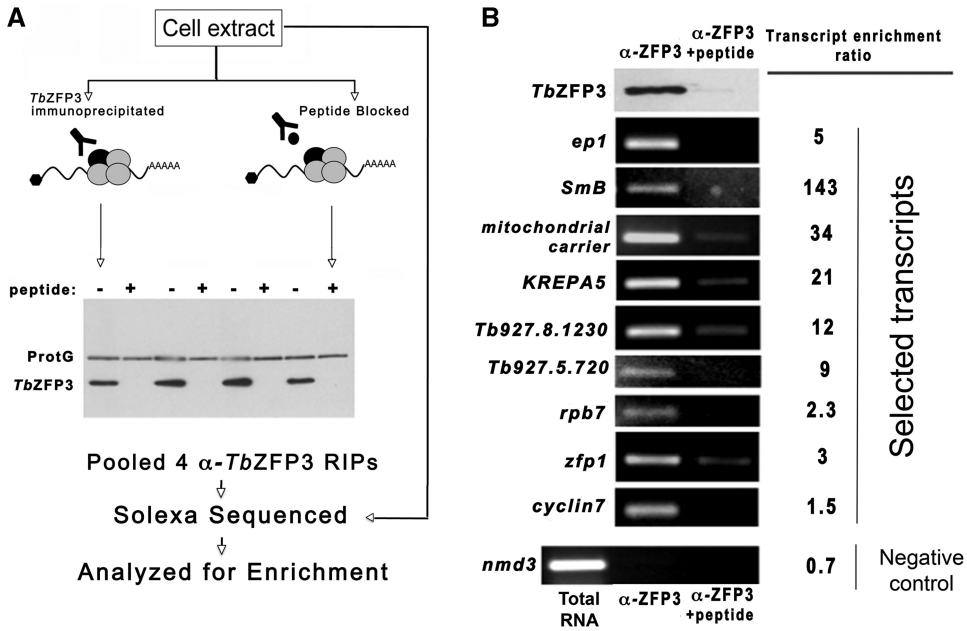

Figure 1. *Tb*ZFP3 selects a specific subset of mRNAs. (**A**) Schematic representation of the selection of *Tb*ZFP3mRNP associating transcripts. Cell extracts were immunoprecipitated (IP) using *Tb*ZFP3-specific anti-peptide antibody, specificity being confirmed using a parallel IP in the presence of blocking peptide. Four separate α-*Tb*ZFP3 IPs were initially validated for their specificity of selection, and the RNA isolated from the pooled material. PolyA+ mRNA was then used to generate cDNA and the population subjected to Illumina Digital Tag sequencing, with the resulting reads compared with unselected procyclic mRNA to identify enriched transcripts in the *Tb*ZFP3-selected pool. (**B**) RT-PCR analysis of *Tb*ZFP3-associated mRNAs isolated in the presence or absence of a *Tb*ZFP3 antibody peptide block. The relative enrichment of the individual transcripts is shown when the combined *ORF* and UTR data sets were considered. A negative control transcript ('*nmd3*') was not enriched in the *Tb*ZFP3 selected mRNA pool, despite its detection in total RNA.

untranslated region. In total, 5790 individual genes could be identified by ORF analysis, with a further 2478 genes being incorporated through 3′-UTR inclusion in the analysis.

To identify the transcripts within the *Tb*ZFP3mRNP, the ratio of each transcript's abundance in the *Tb*ZFP3-selected pool versus the unselected mRNA population was analysed (Supplementary Figure S1). Validating the specificity of the approach used, *EP1* and *GPEET procyclin* transcripts were both enriched in the selected mRNA pools (4.94- and 5.06-fold, respectively; Supplementary Table S1) whereas *EP2 procyclin* was specifically underrepresented (0.593-fold enrichment) and *EP3* was absent (Supplementary Figure S1). This result matched our previous qRT-PCR and regulatory analysis and provided important positive and negative controls that validate this strategy (22). In total, 179 mRNAs were identified as being enriched in the *Tb*ZFP3mRNP pool at least 5-fold, the 48 transcripts showing at least 10-fold enrichment being listed in Table 1, with the full data set being included in Supplementary Table S1. To verify the enrichment of the identified transcripts, independent *Tb*ZFP3 RIP assays were performed and the relative abundance of nine selected mRNAs were analysed by PCR and qRT-PCR. Figure 1B demonstrates that each of the selected transcripts was specifically enriched in the precipitated material and absent or reduced in the peptide-blocked material. A control transcript [*Tb*927.7.970, encoding *Tb*NMD3 (35)], which was not enriched in the *Tb*ZFP3 co-selected material (0.72-fold enrichment), was barely detectable in the selected sample, despite being readily detectable in unselected total mRNA (Figure 1B, 'Negative control'). Independent quantitative analysis by qRT-PCR further validated the selection, demonstrating enrichment of between 5- and 1000-fold with respect to peptide-blocked material for 10 target transcripts (Supplementary Figure S2).

Transcripts selected by the TbZFP3mRNP are also regulated by ectopic over expression of TbZFP3. Having identified that *Tb*ZFP3 RIP co-selected a cohort of diverse transcripts, we tested whether it could regulate these transcripts to validate the relevance of their selection. Hence, we examined the levels of a subset of the selected transcripts in cell lines capable of the ectopic overexpression of *Tb*ZFP3, or a mutant of this protein with the CCCH domain disrupted by point mutation. Further to this, reporter gene assays were used to evaluate the contribution of the 3′-UTR sequences to their regulation, and to determine control at the protein level.

Initially the abundance of three of the *Tb*ZFP3-selected transcripts (*Tb*927.2.4540, '*SmB*', 143-fold enriched; *Tb*09.211.1695, '*SmF*', 16.7-fold enriched; *Tb*10.26.0740, '*Rbp23*', 26.7-fold enriched), and three control transcripts that were not selected ('*TbNMD3*', Tb927.7.970, 0.7-fold enriched; *Tb*927.6.4340, '*TbSm15K*', 0.1-fold enriched; *Tb*09.211.0630, '*actin*', 0.26-fold enriched), were examined in response to overexpression of either a tagged copy of *Tb*ZFP3 (ZFP3-TY), or a mutant in which the predicted RNA binding domain was disrupted

Table 1. Transcripts exhibiting 10-fold or greater enrichment after selection by *Tb*ZFP3 RNA-immunoprecipitation with respect to their tag frequency in unselected RNA

Gene ID	Number of hits		Hit ratio	Expression	Product
	PCF	*Tb*ZFP3mRNP	*Tb*ZFP3mRNP/PCF	ST/SL ratio	
Tb927.2.4540	4	573	143.3	4.29	SmB
Tb927.10.190	59	2702	45.80	14.0	40S ribosomal protein S6
Tb927.6.2820	54	2142	39.67	1.82	Hypothetical protein, conserved
Tb11.03.0870	23	793	34.48	1.97	Mitochondrial carrier protein
Tb927.10.4790	56	1876	33.50	1.70	Hypothetical protein, conserved
Tb927.2.6320	8	259	32.38	1.16	Adenosine transporter 2
Tb927.10.11270	106	2827	26.67	2.55	RNA-binding protein, RBP23
Tb927.2.1260	2	49	24.50	0.43	ESAG, pseudogene
Tb09.211.4513	997	23 683	23.75	0.45	KMP-11
Tb09.160.0420	23	543	23.61	3.09	Hypothetical protein, unlikely
Tb927.10.2170	3	67	22.33	2.33	Hypothetical protein
Tb11.02.0960	1	21	21.00	5.09	Endosomal integral membrane, putative
Tb927.8.680	40	833	20.83	3.31	KREPA5
Tb927.8.6010	860	17 808	20.71	2.55	Predicted transmembrane protein
Tb927.8.550	63	1272	20.19	1.60	Peptide methionine sulphoxide reductase, putative
Tb927.6.700	185	3626	19.60	3.66	Alanyl-tRNA synthetase, putative
Tb927.1.100	7	137	19.57	1.13	RNA polymerase (pseudogene), putative
Tb927.3.5410	767	14 517	18.93	3.24	Hypothetical protein, conserved
Tb09.160.4200	2	36	18.00	1.21	60S acidic ribosomal protein, putative
Tb927.10.3340	124	2105	16.98	0.72	Protein kinase, putative
Tb09.211.2650	237	3979	16.79	0.62	60S ribosomal protein L27a
Tb09.211.1695	231	3853	16.68	1.06	SmF
Tb927.1.2050	17	280	16.47	1.35	Hypothetical protein, unlikely
Tb927.10.3840	4434	68 803	15.52	9.62	60S ribosomal protein L18a, putative
Tb927.10.7740	232	3441	14.83	3.08	Transport prot Sec23A, putative
Tb927.10.14750	18	266	14.78	2.54	Fibrillarin, putative
Tb11.01.0210	72	1061	14.74	1.02	Hypothetical protein
Tb927.10.4560	2557	36 732	14.37	2.34	Elongation factor 2
Tb927.10.7100	167	2372	14.20	1.18	Delta-6 fatty acid desaturase, putative
Tb11.47.0016	336	4725	14.06	1.60	Hypothetical protein, conserved
Tb927.10.6400	292	3646	12.49	2.47	Chaperonin Hsp60
Tb927.8.1230	722	8891	12.31	0.44	Hypothetical protein, conserved
Tb11.01.2060	1	12	12.00	0.47	Hypothetical protein, conserved
Tb11.01.6800	37	441	11.92	1.19	Acyltransferase protein, putative
Tb927.5.4790	4	47	11.75	0.64	VSG, pseudogene, degenerate
Tb927.4.1070	46	528	11.48	1.45	50S ribo prot L13, putative
Tb927.2.2340	7	80	11.43	1.82	Hypothetical protein, conserved
Tb927.1.4830	89	1016	11.42	0.71	Phospholipase A1
Tb11.02.5350	59	666	11.29	3.56	ZC3H43, hypothetical zn finger protein
Tb927.3.5090	128	1443	11.27	3.13	Tryparedoxin, putative
Tb09.160.0465	12 281	133 896	10.90	2.85	Hypothetical protein, conserved
Tb927.4.1790	148	1609	10.87	3.14	Ribosomal protein L3, putative
Tb927.5.460	163	1727	10.60	2.59	Hypothetical protein, conserved
Tb09.160.5590	219	2313	10.56	1.80	60S ribosomal protein L11, putative
Tb927.1.4970	11	116	10.55	4.18	Hypothetical protein
Tb927.7.3410	274	2817	10.28	2.66	Centrin, putative
Tb927.4.1300	4044	40 797	10.09	0.67	Hypothetical protein, conserved
Tb927.4.3880	2	20	10.00	1.92	GRESAG4 R adenylate cyclase

The relative expression of each transcript in pleomorphic slender and stumpy forms is also shown. Two very highly expressed 'Hypothetical conserved' transcripts (Tb09.160.0465, Tb927.4.1300) are present in the table, but no function is evident for these.

by a point mutation (ZFP3 CCAH-TY) (36). In each cell line the tetracycline-inducible expression of the ectopic proteins was confirmed by western blotting using the TY1-specific antibody, BB2 (37) (Figure 2A). Significantly, over expression of *Tb*ZFP3-TY generated an increased abundance of each of the associating transcripts as detected by northern blotting (Figure 2B; compare lanes 3 and 4). This upregulation averaged ~10-fold when these and eight additional selected transcripts were quantitated by qRT-PCR (Supplementary Figures S3 and S4). In contrast, wild-type procyclic forms showed no tet-induced changes in target transcript abundance (Figure 2B, lanes 1 and 2) and mutation within the CCCH domain of *Tb*ZFP3 prevented regulation, although in this case the level of expressed protein was less (Figure 2A and B; lanes 5 and 6). Importantly, the upregulation of associating transcripts by *Tb*ZFP3 levels was specific as none of the negative control transcripts were elevated (Figure 2C). These results indicate that increased levels of *Tb*ZFP3 upregulate levels of associating transcripts, this requiring an intact CCCH domain.

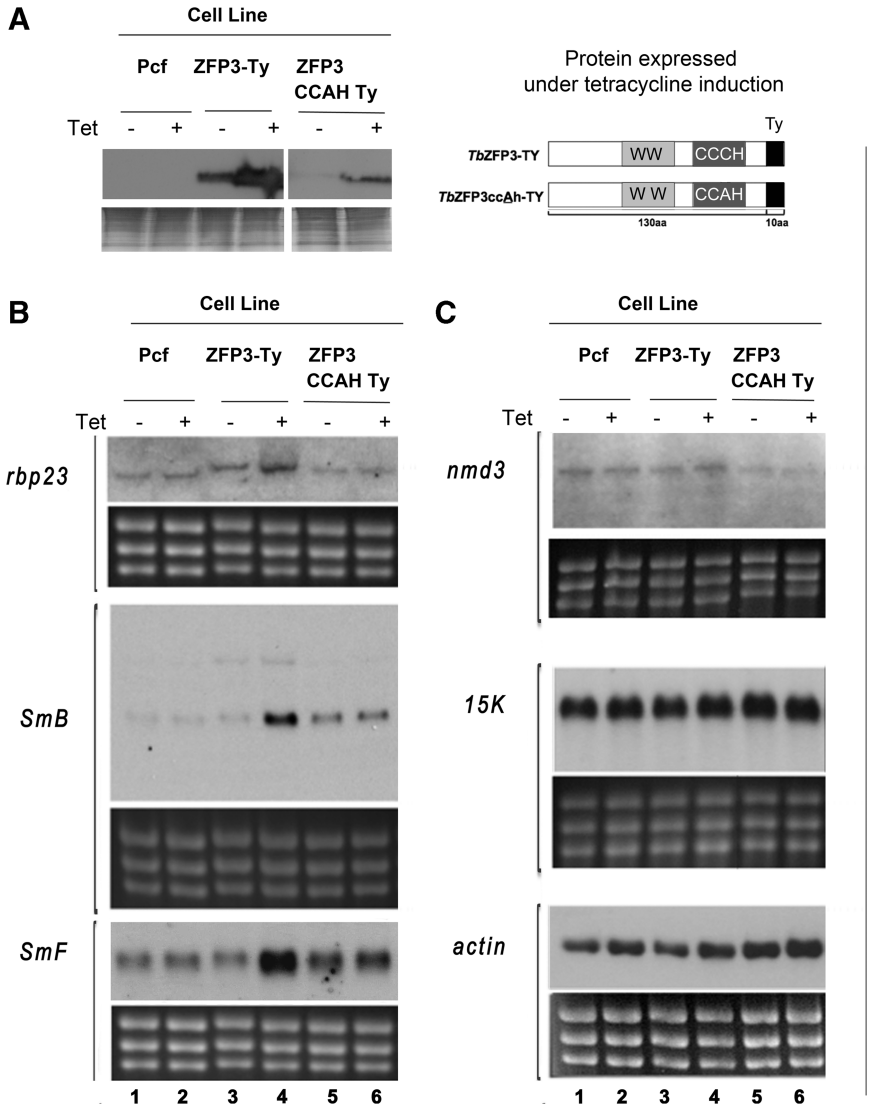

Figure 2. *Tb*ZFP3-selected transcripts are specifically upregulated by ectopic *Tb*ZFP3, dependent upon the predicted RNA binding domain. (**A**) Expression of ectopic *Tb*ZFP3-TY and *Tb*ZFP3ccAh-TY in cells induced, or not, with tetracycline. The expressed proteins were detected using BB2 antibody ('anti-TY'), specific for the incorporated Ty1 epitope tag in each protein. A schematic representation of the ectopically expressed proteins is shown to the right. (**B**) Detection of the transcript abundance for three target mRNAs selected by *Tb*ZFP3. The abundance of each transcript is shown when the following cells lines were induced with tetracycline: lanes 1 and 2, parental procyclic forms; lanes 3 and 4, ectopic *Tb*ZFP3-TY; lanes 5 and 6, ectopic *Tb*ZFP3 CCAH-TY. (**C**) As in B, for three negative control transcripts that are not selected by *Tb*ZFP3. Note that a small increase in *actin* transcript was observed upon tetracycline induction in this experiment, but that this was not clearly reproducible between independent experiments.

We have previously demonstrated that *Tb*ZFP3 regulates *EP1 procyclin* gene expression through regulatory elements within the 3′-UTR (22). In order to assess whether the 3′-UTR's of other target transcripts were implicit in *Tb*ZFP3-dependent regulation, the intergenic region spanning from the stop codon of either *Rbp23*, *SmB* or *GrpE* to the borders of the next gene were inserted into constitutive CAT reporter constructs (Figure 3A). These CAT-*target*UTR reporter constructs were then stably transfected into multiple cell lines (Figure 3A) including the parental control and lines that ectopically express either native *Tb*ZFP3 ('ZFP3 No Tag'), *Tb*ZFP3-TY or the mutant protein *Tb*ZFP3 CCAH-TY. The resulting lines (summarized in Figure 3A) were then tested for tet-inducible ectopic *Tb*ZFP3 expression using antibodies detecting either *Tb*ZFP3 or the TY tag (Figure 3B). This revealed that all reporter cell lines exhibited inducible *Tb*ZFP3 expression, except for the 'CAT-*SmB*' reporter in the 'ZFP3 No Tag' cell line, where ectopic expression was considerably leaky, such that equivalent ectopic *Tb*ZFP3 expression was observed whether induced or not.

Figure 4A shows the effect of *Tb*ZFP3 ectopic overexpression upon these *CAT* reporters. For both *Rbp23 and SmB*, the 3′-UTR sequence was sufficient to generate an inducible increase in *CAT* mRNA in response to ectopic *Tb*ZFP3-TY expression (Figure 4A, left panel; lanes 2 and 3 for *Rbp23*, and lanes 4 and 5

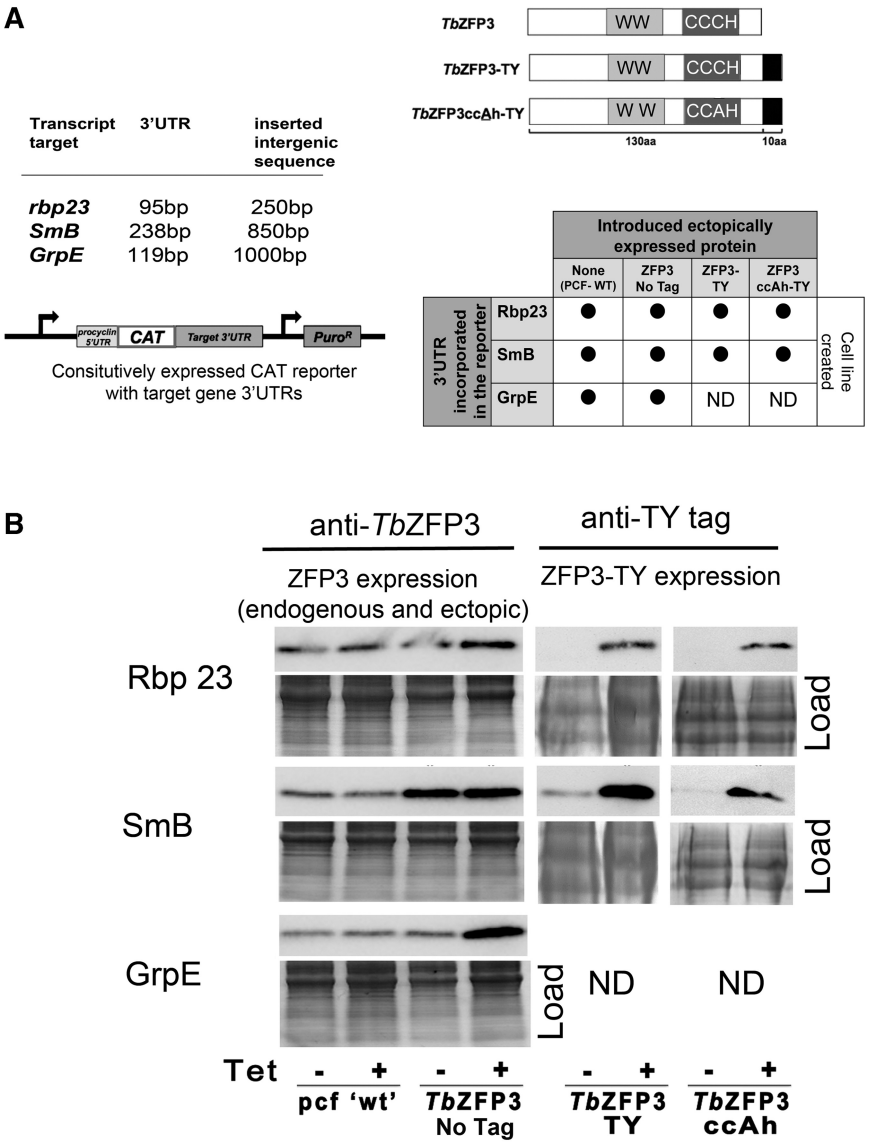

Figure 3. Cell lines created to investigate the regulation by *Tb*ZFP3 of target mRNA 3′-UTRs. (**A**) Chloramphenicol acetyltransferase (CAT) Reporter strategy. Intergenic sequences downstream of *Rbp23, SmB* and *GrpE* coding regions were inserted individually into a reporter construct adjacent to the CAT coding region. In each case, the inserted sequence length is indicated in the Table in Panel A, as is the predicted endogenous 3′-UTR length for each transcript. Each reporter was transfected into parental 'wild-type' procyclic form cells and those capable of the inducible ectopic expression of untagged *Tb*ZFP3 (ZFP3-No Tag), *Tb*ZFP3-TY or ZFP3 CCAH-TY. The respective cells lines generated in each case are summarized in the chart. ND = not done. (**B**) Western blots of inducible ectopic *Tb*ZFP3 expression in lines transfected with the respective CAT reporter constructs. Note that for the *CAT-SmB*/ZFP3-No tag cell line, the level of ectopic *Tb*ZFP3 expression was equivalent in uninduced and induced cells. Hence, for this cell line expression comparisons of the CAT reporter were made with the parental *CAT-SmB* transfected line. Relative loading is indicated by the Coomassie stained gel images in each case. ND = Not done.

for *SmB*), whereas the *Tb*ZFP3-CCAH-TY mutant generated no effect (Figure 4A, left panel, lanes 6–9). Similarly, untagged ectopic *Tb*ZFP3 elevated the *Rbp23* reporter transcript levels upon induction (Figure 4A, right panel, lanes 7 and 8). For the *SmB* reporter, the leakiness of ectopic *Tb*ZFP3 protein in this cell line (Figure 3B) required comparison with the parental cell line transfected with the same *SmB* reporter (Figure 4A, right panel, compare lanes 3 and 4 with lanes 9 and 10). This analysis again indicated enhanced abundance of the *SmB* reporter mRNA associated with elevated *Tb*ZFP3 expression. We therefore conclude that the 3′-UTR's of both *Rbp23* and *SmB* contain *Tb*ZFP3-responsive regulatory elements.

Surprisingly, the *GrpE* reporter mRNA was not altered by ectopic *Tb*ZFP3 expression (Figure 4A, right panel, compare lanes 5 and 6 with lanes 11 and 12). This lack of response contrasts with the clear effect of *Tb*ZFP3 ectopic expression upon the endogenous *GrpE* transcript in the same cell lines (Supplementary Figure S4). This demonstrated that sequences within 1-kb downstream of the *GrpE* coding region were not responsive to elevated *Tb*ZFP3 levels, indicating that the regulation of *GrpE* transcript is dependent upon elements outside those

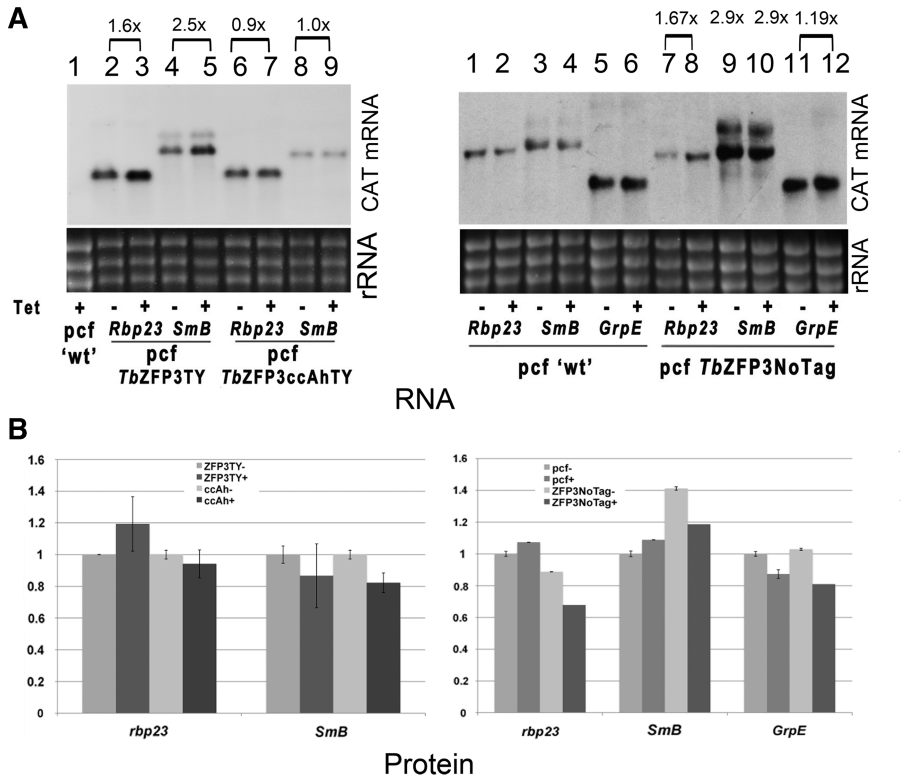

Figure 4. The 3′-UTR's of SmB and Rbp23 are sufficient for regulation by *Tb*ZFP3. (**A**) Ectopic overexpression of *Tb*ZFP3-TY (left panel) or untagged *Tb*ZFP3 (right panel) increases levels of *CAT-Rbp23* and *CAT-SmB,* matching the effect on the endogenous mRNAs. This upregulation is dependent upon the CCCH predicted RNA-binding domain of *Tb*ZFP3. In contrast, the *CAT-GrpE* transcript was not significantly changed in response to ectopic *Tb*ZFP3 expression, unlike endogenous *GrpE* mRNA (Supplementary Figure S3). In each case, a northern blot is shown detecting *CAT* mRNA. Relative loading is indicated by EtBr stained rRNA. The relative fold increases for each reporter mRNA, based on their chemifluorescent signal, are highlighted above the lane numbers used for comparison, these being derived from an independent experiment. For the *CAT-SmB* reporter in the '*Tb*ZFP3 No Tag' line, the leakiness of ectopic protein expression necessitated comparison with the PCF control line (lanes 3 and 4). (**B**) Corresponding CAT protein levels are not significantly upregulated in response to ectopic *Tb*ZFP3-TY (left-hand panel) or *Tb*ZFP3 (right-hand panel) overexpression. In each case CAT protein levels were determined by CAT-ELISA assay and normalized to the uninduced ZFP3-TY line (left-hand panel) or PCF (without tetracycline) containing the reporter constructs, but no ectopic *Tb*ZFP3.

included in the reporter construct, i.e. the *GrpE* 5′-UTR or *ORF*.

Having demonstrated that ectopic *Tb*ZFP3 overexpression increases *CAT* reporter mRNA levels controlled by both *Rbp23* or *SmB* 3′-UTR's, we investigated whether this upregulation translates to the CAT protein level via quantitative CAT-ELISA assay (Figure 4B). However, in neither case was inducible elevation of the reporter protein observed upon over expression of either *Tb*ZFP3-TY (Figure 4B, left hand panel) or untagged *Tb*ZFP3 (Figure 4B, right hand panel). This revealed that, unlike *EP1 procyclin* (22), ectopic *Tb*ZFP3 expression specifically increased target mRNA abundance but generated no significant change in protein levels for the examined transcript targets. This suggests that translation is restricted despite the elevated mRNA abundance for these reporter mRNAs.

To determine the basis of the increased abundance of the selected mRNAs, the decay of *SmB* (*Tb*ZFP3-selected) and *TbSm15K* (non-selected) transcripts were assayed after actinomycinD treatment of cells induced, or not, to overexpress *Tb*ZFP3. Initially, we confirmed the inducible overexpression of *Tb*ZFP3 in the cells (Supplementary Figure S5) and examined the relative levels of *SmB, Rbp23* and *GrpE* versus the non-target negative controls *TbSM15k* and *SmB* and actin. Semi quantitative northern blot data confirmed the specific upregulation for the selected transcripts (data not shown). Thereafter, the decay of *TbSm15k* and *SmB* mRNA was analysed in induced and uninduced cells by qRT-PCR (Figure 5A and B). This revealed that the decay of *TbSm15K* was unaffected by the overexpression of *Tb*ZFP3 ($F_1 = 0.10$, $P = 0.759$; Figure 5A) whereas *SmB* exhibited reduced decay rates in the *Tb*ZFP3-induced line compared to the uninduced cells ($F_1 = 5.62$, $P = 0.050$, Figure 5B), doubling the *SmB* transcript half-life from 40 to 80 min. Semi-quantitative northern blot data confirmed this response for the *GrpE* and *Rbp23* transcripts (data not shown). These assays indicated that *Tb*ZFP3-selected transcripts are stabilized by *Tb*ZFP3 overexpression, providing an explanation for the elevated abundance of target transcripts in steady state mRNA.

TbZFP3mRNP transcripts are enriched in the parasite transmission stage. Having confirmed the specificity and regulation of *Tb*ZFP3-associated transcripts, we

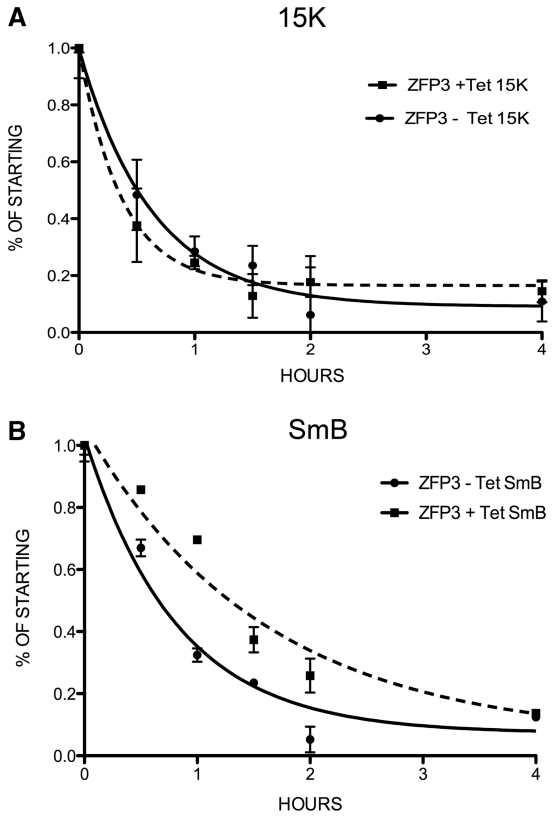

Figure 5. *Tb*ZFP3 stabilizes associating transcripts. (**A**) mRNA quantitation derived from triplicate qRT-PCR assays of *TbSm15K* mRNA (not selected by *Tb*ZFP3) at time points after treatment with actinomycin D. Samples were derived from cells induced, or not, to ectopically express *Tb*ZFP3, with values expressed as a percentage of the starting abundance. Using a GLM with % of starting transcript as the response variable, the presence/absence of TET was not a significant factor ($F_1 = 0.10$, $P = 0.759$). (**B**) As in A, but assayed for the abundance of the *SmB* (selected by *Tb*ZFP3). The steady state abundance of *SmB* is enhanced upon *Tb*ZFP3 ectopic expression, hence comparisons between induced and uninduced samples are expressed as a percentage of the abundance at time = 0 h in each case. In this case, the presence/absence of TET was a significant factor ($F_1 = 5.62$, $P = 0.050$).

investigated whether they exhibited evidence of functional co-ordination or co-expression during the parasite's lifecycle. To achieve this, the relative abundance of each *Tb*ZFP3-RIP selected transcript was compared between different developmental forms of the parasite, namely bloodstream 'slender' forms, 'intermediate' forms and 'stumpy' forms, and cultured tsetse fly midgut procyclic forms (Supplementary Figure S6). The bloodstream forms were derived from Day 3 (slender), Day 5 (intermediate) and Day 7 (stumpy) of a mouse infection, and replicate samples were generated from two independently isolated *T. brucei* strains (AnTAT1.1 or EATRO), thereby ensuring that any identified developmental profile was consistent between strains. These RNAs were then subject to Illumina Digital Tag gene expression analysis to derive the abundance of individual transcripts at each developmental stage and thereafter those transcripts selected by *Tb*ZFP3 analysed for their relative expression profile. Figure 6 and Supplementary Figure S6 show the relative expression of the top 100 (>7-fold enriched; Figure 6A and B), 200 (>4.75-fold enriched), 300 (>3.5-fold enriched), 400 (>3-fold enriched) and 500 (>2.5-fold enriched) *Tb*ZFP3-selected transcripts compared to the total pool of all transcripts in slender or stumpy forms. Strikingly, the selected transcripts were significantly overrepresented in the stumpy developmental form in both strains of *T. brucei* tested. Furthermore, the extent of representation directly correlated with the relative enrichment of the transcripts in the *Tb*ZFP3-selected material (Figure 6B). Hence, the top 100 enriched transcripts in the *Tb*ZFP3 selected pool showed significantly elevated expression in the stumpy derived mRNA pool ($\chi^2 = 10.8$, df = 2, $P < 0.005$), with progressively less evidence for stumpy-enriched expression when the top 200, top 300 and top 400 *Tb*ZFP3-selected transcripts were considered (Figure 6B). Correspondingly, the *Tb*ZFP3-RIP selected transcripts were enriched in 'intermediate forms', albeit less dramatically than in stumpy forms, and were underrepresented in the mRNA pool enriched in slender forms (Supplementary Figure S6 and Figure 6).

To investigate whether the enriched genes showed any functional co-ordination, the Gene Ontology (GO) of the top 100 enriched mRNAs was investigated using the GOstat software package (39). This identified GO groups that were over-represented in the selected list compared to the overall list of *T. brucei* GO annotated genes (www.geneDb.org; 8 December 2011 update). Using a stringent *P*-value cutoff of 0.01, this analysis revealed five GO groups that showed significant over-abundance in the top 100 *Tb*ZFP3mRNP transcripts compared to a randomized GO annotated *T. brucei* 100 gene set (Figure 7). These groups were 'Ribonucleoprotein Complex' ($P = 0.002$), 'Macromolecule Biosynthetic Complex' ($P = 0.003$), 'Translation' ($P = 0.009$), 'Lipase Activity' ($P = 0.01$) and 'Cytosolic Large Ribosomal Subunit' ($P = 0.01$). These GO identities make up the majority of the *Tb*ZFP3mRNP pool and are also enriched in stumpy enriched mRNA pool, contrasting with the total transcriptome population (Figure 7).

Combined, these analyses demonstrated that *Tb*ZFP3 preferentially co-associated with transmission stage enriched mRNAs, invoking the presence of a novel developmental regulon. Moreover, the selected mRNAs were enriched for molecules likely to be necessary as parasites prepare for the extensive changes in gene expression and protein synthesis upon vector uptake.

TbZFP3 associates into Procyclic form cytoplasmic granules upon serum starvation but not in stumpy forms. Having determined that *Tb*ZFP3 preferentially associates with transcripts enriched in stumpy forms, we investigated whether this was co-ordinated through any higher order mRNP structure. In eukaryotic cells, transcripts can be stabilized through their association with cytoplasmic granules, which can provide storage sites under conditions of stress or nutritional starvation. Several predicted RNA binding proteins in kinetoplastids localize into cytoplasmic foci upon serum starvation, or in

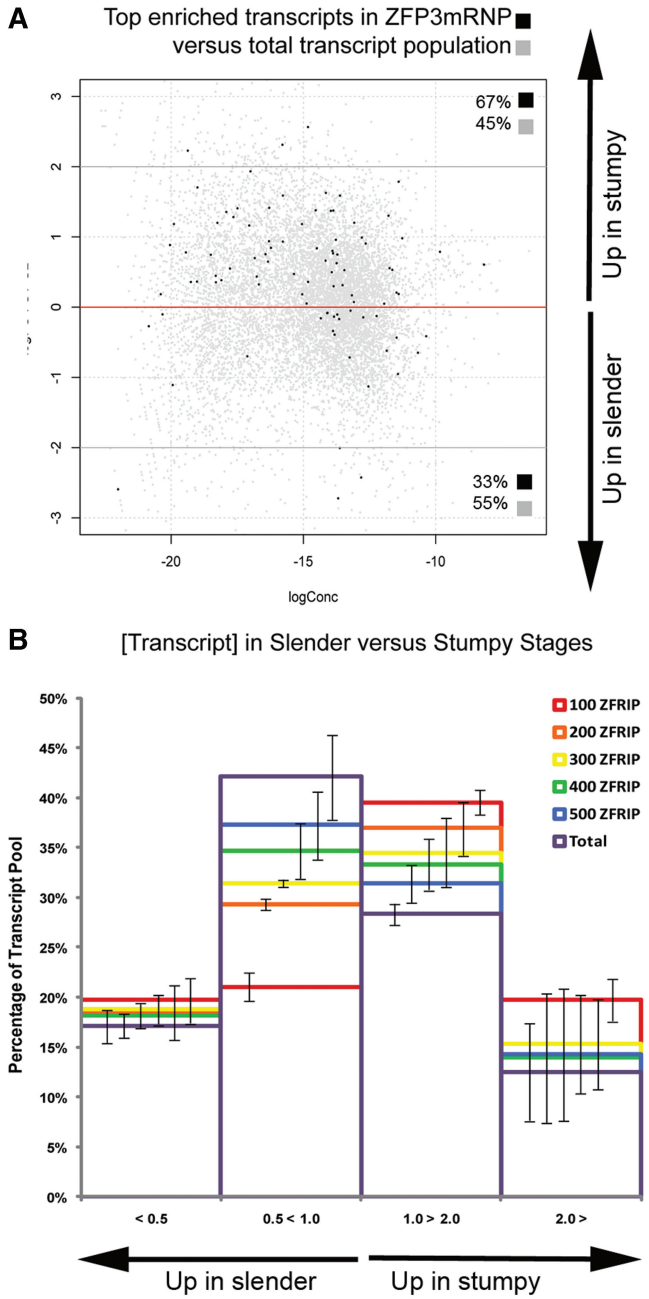

Figure 6. *Tb*ZFP3-associated transcripts are enriched in stumpy forms. (**A**) EdgeR representation of the ratio of transcript concentrations in stumpy (ST) versus slender (SL) stage parasites of the total transcript pool (light grey) versus the 100 most enriched transcripts in the *Tb*ZFP3-selected pool (black). The distribution of the top 100 enriched transcripts in the *Tb*ZFP3-RIP selected material with respect to their relative expression in slender forms (<0) or stumpy forms (>0) is shown. The selected transcripts are predominantly (67%) in the stumpy-enriched cohort. (**B**) Relative expression of the top 100, top 200, top 300, top 400 and top 500 ('100 ZFRIP', etc.) *Tb*ZFP3-RIP selected transcripts ranked in order of their enrichment in slender or stumpy forms with respect to unselected material. The selected transcripts are more predominant amongst those transcripts enriched in stumpy forms ('Up in stumpy'), and less predominant in those transcripts enriched in slender forms ('Up in slender'). The relative enrichment after *Tb*ZFP3-RIP correlates with the extent of enrichment in stumpy forms (one-way ANOVA analysis, $F_{11} = 6.18$, $P = 0.023$) although the correlation was less significant for transcripts >2-fold enriched in stumpy forms ($F_{11} = 3.27$, $P = 0.091$). This inversely correlates with the extent of enrichment in slender forms ($F_{11} = 6.18$, $P = 0.023$) and for transcripts >2-fold enriched in slender forms compared to stumpy ($F_{11} = 14.93$, $P = 0.002$). *Post hoc* Tukey's tests indicate that the Top100, Top200 and Total categories are the important factors in those ANOVA that are significant. Data represent the analysis of two independent pleomorphic slender and stumpy samples derived from different strains of *T. brucei*. Error bars for each transcript group are shown.

response to heat stress (38,39). Although several classes of mRNP granule have been observed in trypanosomatids, the molecule SCD6 is believed to be diagnostic for cytoplasmic P body granules (38). To determine whether *Tb*ZFP3 could associate with these mRNP granules, the cellular location of this molecule was established by expressing a SCD6-eYFP fusion in procyclic forms [a kind gift of Dr Mark Carrington; (38)]. In SDM79 procyclic cell medium, both *Tb*ZFP3 and SCD6-YFP were dispersed throughout the cell cytoplasm (Figure 8A). However, when incubated in serum-free phosphate buffered saline for 2 h, both proteins coalesced into discrete cytoplasmic foci that colocalized (74% correspondence) (Figure 8B). Unlike Scd6, however, this redistribution of *Tb*ZFP3 was not observed in response to heat shock (42°C; data not shown), indicative of a starvation-specific rather than broader stress-induced relocalization. This indicates that *Tb*ZFP3 associates with P body granules specifically upon serum starvation in procyclic forms.

To probe the functional significance of this P-body association, we examined *Tb*ZFP3 localization in stumpy forms, when cells might be predicted to be under stresses analogous to PBS starvation in procyclic forms. Although reagents are not available to visualize SCD6 in pleomorphic bloodstream forms, stumpy cells harvested from a mouse infection and purified from host blood by DEAE chromatography were stained for *Tb*ZFP3 to identify granule-like structures similar to those seen in procyclic forms. The same cells were also analysed by *in situ* hybridization, to visualize *EP procyclin* transcript localization. This *EP*-specific mRNA probe detects all three *EP procyclin* mRNA isoforms, including *EP2* and *EP3*, which do not associate with *Tb*ZFP3. However, we anticipated that *EP1 procyclin* mRNA and *Tb*ZFP3 protein might colocalize into detectable P body-like structures. Figure 8C shows that in stumpy forms, *Tb*ZFP3 exhibited a diffuse but punctate cytoplasmic staining, with some local concentration in areas of the cytoplasm. Similarly, *procyclin* transcripts visualized with an anti-sense probe were diffusely located, whereas a sense probe generated no signal (Figure 8D), supporting the specificity of the detected signal. However, large, discrete foci as induced by serum starvation in procyclic cells were not evident in stumpy cells, despite some local concentration of both signals (arrowed in Figure 7C). Hence, *Tb*ZFP3 can associate with P bodies upon serum starvation in procyclic forms, but analogous structures are not obvious in transmissible stumpy forms.

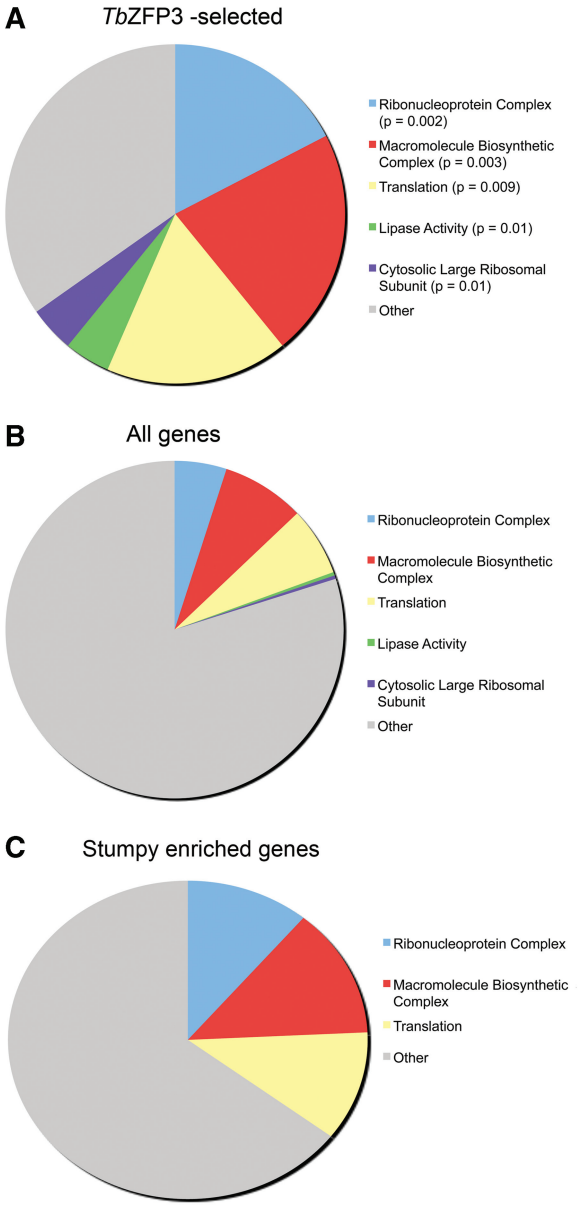

Figure 7. GO analysis of the *Tb*ZFP3mRNP associating transcripts. Representation of the GO terms enriched in the *Tb*ZFP3-RIP selected transcripts (**A**) versus the relative abundance of the same GO classes in the total (**B**) or stumpy enriched (**C**) gene set. *P*-values were calculated using the GOStat package (40).

DISCUSSION

Although there are large numbers of predicted RNA binding proteins encoded in the genomes of kinetoplastid parasites (36,41), only in a few instances have target transcripts been identified. Of these, the best characterized are (i) the cell cycle box binding proteins, CSBPA and B, which recognize a conserved octamer sequence in the UTR's of cell cycle-regulated transcripts (15), (ii) Puf 9 (42), which also has a putative role with cell cycle-regulated mRNAs, (iii) DRBD3, an RGG domain protein that appears to associate with the mRNAs of membrane proteins (43) and (iv) the small CCCH protein family comprising *Tb*ZFP1, *Tb*ZFP2 and *Tb*ZFP3. Of the latter, each is less than 140 amino acids and co-associate in procyclic forms, aided by complementary protein interaction domains (31). The *Tb*ZFP proteins have each been implicated in regulating developmental processes, namely kinetoplast repositioning [*Tb*ZFP1; (26)] or the efficiency of differentiation as monitored by the expression of the Procyclin surface proteins and morphology [*Tb*ZFP2, *Tb*ZFP3 (24,31)]. In the case of the *Tb*ZFP3mRNP, the interaction with *procyclin* mRNA was found to be direct. Specifically, the *Tb*ZFP3mRNP associated with the Loop II element of the *EP1 procyclin* 3′-UTR elevating the levels of EP1 Procyclin protein at the cell surface (22), this being dependent on the integrity of the *Tb*ZFP3 CCCH domain. Here, we have carried out a global survey of the mRNAs that co-associate with the *Tb*ZFP3mRNP. This has revealed, firstly, that the selected mRNAs were stabilized by *Tb*ZFP3 and, secondly, that the selected transcripts were predominantly more abundant in the transmission stage of trypanosomes, stumpy forms. This implicates *Tb*ZFP3 mRNP as a *trans*-acting factor defining a developmental regulon in these parasites.

The approach used to identify mRNAs that associate with the *Tb*ZFP3 mRNP involved co-immunoprecipitation of mRNAs by an anti-peptide antibody specific for *Tb*ZFP3 (22). The selected transcripts were then identified by their relative enrichment with respect to unselected mRNA, this being determined quantitatively at a global level by use of Illumina Digital-tag expression analysis. This approach offers a number of benefits. Firstly, by use of an anti-peptide antibody, a blocking peptide control could be incorporated into the selection regime, ensuring that interactions were specific for the target protein, *Tb*ZFP3. Secondly, by use of an antibody to the endogenous protein we could ensure that the physiological stoichiometry of mRNA–mRNP interactions in the cell was preserved and avoid the need for affinity tags to be incorporated into the protein ligand. The latter is an important consideration since we have observed that incorporation of a 10 amino acids TY tag into the C-terminus of *Tb*ZFP3 alters selection of *procyclin* isoform mRNAs (our unpublished data). Finally, by use of high-throughput Digital-tag transcriptome analysis, we could accurately identify and quantify selected transcripts, exploiting the available *ORF* and the 3′-UTR data generated by RNA-seq analysis of trypanosome life cycle stages (10).

Analysis of the developmental expression profile of the transcripts co-associated with *Tb*ZFP3 revealed an enrichment of mRNAs whose expression is elevated in stumpy forms, which are poised for development when taken up in a tsetse bloodmeal (44,45). Indeed, there was a strong correlation between the extent of enrichment after *Tb*ZFP3 RIP and stumpy-enriched expression, this trend being observed in two independently isolated parasite lines capable of transmission. This observation is consistent with the established roles of *Tb*ZFP proteins in bloodstream to procyclic form differentiation (22,36,25,26) and the enrichment of CCCH proteins in differentiation events recently observed by high-throughput RNAi

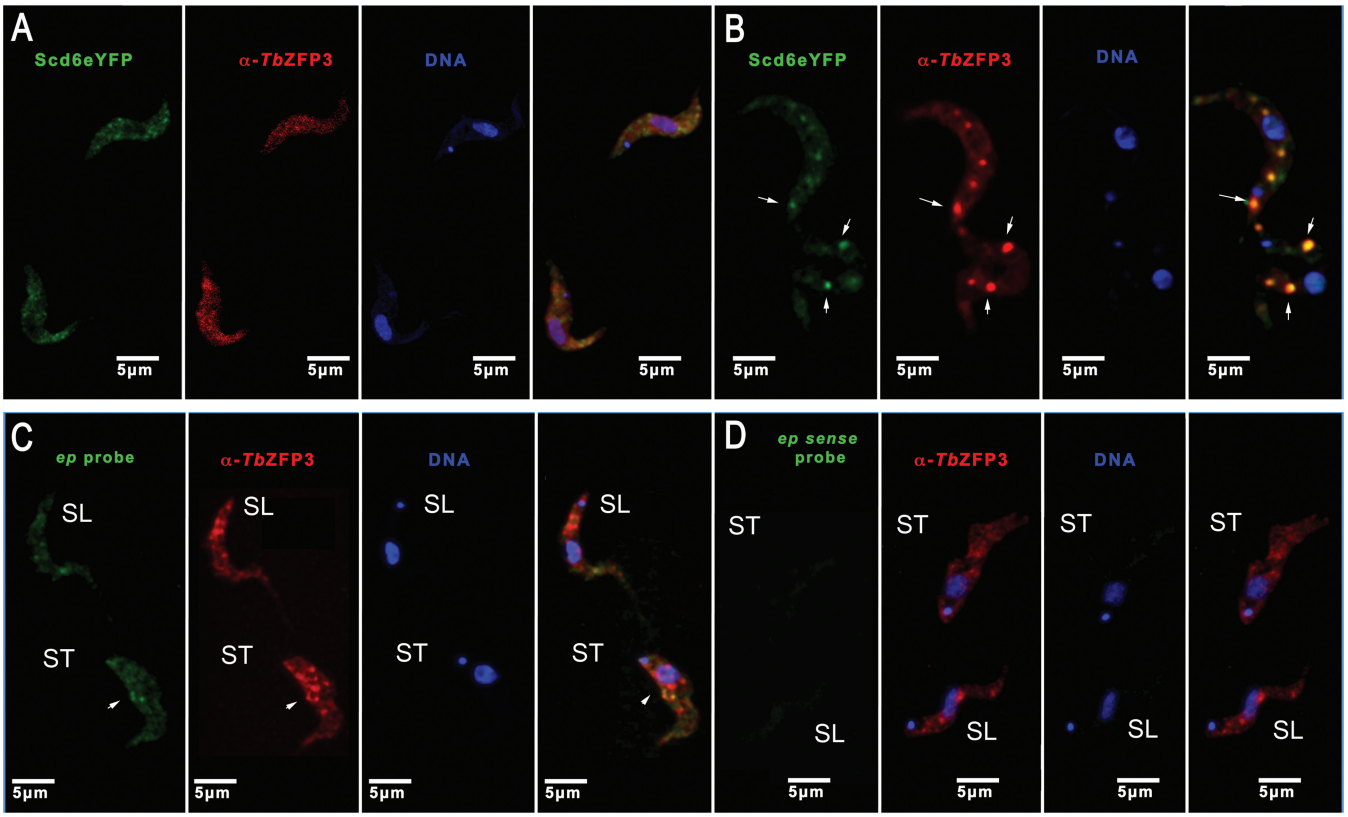

Figure 8. *Tb*ZFP3 association in cytoplasmic P granules. Panels (**A**) and (**B**) show Procyclic form cells engineered to express *Tb*SCD6-YFP were incubated in SDM-79 (A) or PBS for 2 h. (B) and then the location of *Tb*SCD6 YFP (green) or *Tb*ZFP3 (red) determined by immunofluorescence. The third panel in each case represents the image derived from DAPI staining to visualize the parasite nucleus and kinetoplast, whereas the fourth panel shows a merge of the *Tb*SCD6-YFP, *Tb*ZFP3 and DAPI panels. After PBS starvation the *Tb*SCD6-YFP and *Tb*ZFP3 signals colocalize in discrete cytoplasmic granules or P Bodies (arrowed). Panels (**C**) and (**D**) show the signal generated when slender (SL) and stumpy (ST) cells are probed for the location of *ep procyclin* mRNA (panel C, anti-sense probe, panel D, sense probe) and *Tb*ZFP3, with DAPI staining of the cells being shown in the third panel and a merge of all staining in the fourth panel. The *Tb*ZFP3 signals do not localize into enlarged P bodies, unlike serum starved procyclic forms. However, some concentration, in particular regions is observed (arrowed).

screening (46). Coupled with the regulation by *Tb*ZFP3 of *EP1 procyclin*, the earliest marker of development to procyclic forms, we propose that the small CCCH proteins are implicit in controlling the changes in gene expression that accompany life-cycle development upon entry to the tsetse fly. Supporting this, *Tb*ZFP3 co-associated with mRNAs functionally linked to gene regulation and new protein synthesis, a profile expected for control of the early events during developmental progression upon entry into the tsetse fly midgut.

Analysis of a subset of transcripts selected by *Tb*ZFP3-RIP revealed that each showed increased abundance after *Tb*ZFP3 levels were elevated by ectopic expression, leading to enhanced mRNA stability. This differs from our earlier observations with *procyclin* mRNAs where *Tb*ZFP3 overexpression did not alter levels of the specifically selected *EP1* and *GPEET* mRNAs (22). Also contrasting with *EP1 procyclin* regulation by *Tb*ZFP3, we did not find evidence for enhanced protein expression of the target mRNAs linked to elevated *Tb*ZFP3 expression. While this evidence may superficially appear to conflict and indicate distinct regulatory functions, in all cases examined *Tb*ZFP3 acts to positively regulate associating targets. Hence, through stabilization it appears *Tb*ZFP3 potentiates but does not ensure the translation of associating transcripts. This model is consistent with global analysis of mRNA and protein levels observed in the related kinetoplastid, *Leishmania donovani*, where mRNA were predicted to be stabilized and directly targeted for translation upon a differentiation signal (47,48).

The regulatory distinctions for different *Tb*ZFP3mRNP target mRNAs may result from several processes. Firstly, different transcript classes may have different rate-limiting steps in their regulatory control. *Procyclin* represents an mRNA that must be exquisitely regulated to prevent the premature appearance of protein on the surface of bloodstream form parasites, where it could provoke a strong immune response against the parasite. Consequently, this transcript must be stringently limited by transcript-specific translational repression and transcript destabilization (21), a restriction counteracted by the

'anti-repressor' effect of the *Tb*ZFP3mRNP association with the Loop II regulatory region of the *procyclin* mRNA 3′-UTR (49). In contrast, for other target transcripts, target mRNA are stabilized by *Tb*ZFP3, whereas protein expression might be restricted by a more general translational control mechanism or other regulatory factors. Secondly, variation in the precise protein composition of *Tb*ZFP3mRNP(s) may exist for different transcripts or transcript classes. In this scenario, *Tb*ZFP3 could associate with distinct protein factors altering mRNP specificities that regulate different target mRNA classes. Finally, we cannot rule out limitations imposed by the reporter system used. Here, two 3′-UTR sequences (for *SmB* and *RBP23*) were found to recapitulate the elevated target mRNA observed when *Tb*ZFP3 was ectopically expressed. However, if other sequences in the 5′-UTR or coding region contribute to gene regulation, matching the scenario for *GrpE* and at least one other transcript (16), then our assays may not fully represent the regulatory consequences of increased *Tb*ZFP3 levels. In all cases, however, other factors in addition to *Tb*ZFP3 must operate to differentially regulate gene expression because unlike our perturbation experiments, the endogenous levels of *Tb*ZFP3 do not dramatically differ between bloodstream and procyclic forms (31). One such factor is likely to be *Tb*ZFP1, which is induced during differentiation and interacts with *Tb*ZFP3 (31). Another is the differential polysome association of *Tb*ZFP3 between life cycle stages (40).

How is the specificity of *Tb*ZFP3 regulation achieved? A bioinformatic analysis of the 3′-UTR sequences of the selected transcripts revealed an enrichment of sequence motifs that distinguish *EP1* and *GPEET* from *EP2* and *EP3* 3′-UTR sequences, supportive of a sequence specific interaction (Supplementary Figure S7). However, this relationship was not simple, and not all enriched transcripts shared the same sequence motifs. This is not surprising as secondary structural features in the target mRNAs may contribute to *Tb*ZFP3 recognition, such structural motifs being difficult to predict and identify by computational means alone. Also, as discussed earlier, *Tb*ZFP3 may associate with different mRNP complexes with different specificities, complicating the identification of conserved motifs among the global cohort of transcripts co-selected with *Tb*ZFP3. Although further analysis of the RNA–protein and protein–protein interactions of this *trans*-acting regulatory factor are necessary, its association into cytoplasmic granules containing *Tb*SCD6 in procyclic forms demonstrate its involvement in the higher order mRNA regulatory complexes within the cell (Figure 8).

To conclude, we have exploited a proven and physiologically relevant strategy to identify the cellular population of mRNAs specifically associated with the small *trans*-acting post-transcriptional regulator, *Tb*ZFP3. In each case, *Tb*ZFP3 was found to act as a positive regulator of gene expression, generating increased stability for associated mRNAs, matching another CCCH class protein in *T. brucei* (49). Strikingly, the selected mRNAs were enriched in transmission stages of the parasite, suggesting the possibility of a developmental post-transcriptional operon directed by *Tb*ZFP3, rather than a more limited functional operon. Moreover, the broad range of enrichment for different transcripts suggests that there will be considerable complexity and nuance in the regulation of specific transcript and transcript groups by CCCH family RNA regulators. By identification of the full complement of *Tb*ZFP3 mRNA targets, these interactions and regulatory events can now be analysed in detail, greatly extending earlier 'one transcript-one regulator' models for gene regulation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1–7.

ACKNOWLEDGEMENTS

We thank S. Kramer and M. Carrington, University of Cambridge, for the gift of the *Tb*SCD6-eYFP cell line and useful discussion. D. Murray of the Parasite Imaging Facility was instrumental in confocal analyses. This manuscript benefited from the comments of A. Schnaufer.

FUNDING

Wellcome Trust (088293, Programme Grant to K.R.M.) (supporting P.B.W, P.C. and K.F.) and Wellcome Trust, The Centre for Immunity, Infection and Evolution (095831, strategic award). Funding for open access charge: Wellcome Trust.

Conflict of interest statement. None declared.

REFERENCES

1. Moore,M.J. (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science*, **309**, 1514–1518.
2. Houseley,J., LaCava,J. and Tollervey,D. (2006) RNA-quality control by the exosome. *Nat. Rev. Mol. Cell. Biol.*, **7**, 529–539.
3. Keene,J.D. (2007) RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.*, **8**, 533–543.
4. Wilkinson,M.F. and Shyu,A.B. (2001) Multifunctional regulatory proteins that control gene expression in both the nucleus and the cytoplasm. *Bioessays*, **23**, 775–787.
5. Clayton,C. and Shapira,M. (2007) Post-transcriptional regulation of gene expression in trypanosomes and *Leishmania*s. *Mol. Biochem. Parasitol.*, **156**, 93–101.
6. Schwede,A., Kramer,S. and Carrington,M. (2011) How do trypanosomes change gene expression in response to the environment? *Protoplasma*, doi:10.1007/s00709-011-0282-5.
7. Siegel,T.N., Hekstra,D.R., Kemp,L.E., Figueiredo,L.M., Lowell,J.E., Fenyo,D., Wang,X., Dewell,S. and Cross,G.A. (2009) Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev.*, **23**, 1063–1076.
8. Berriman,M., Ghedin,E., Hertz-Fowler,C., Blandin,G., Renauld,H., Bartholomeu,D.C., Lennard,N.J., Caler,E.,

Hamlin,N.E., Haas,B. *et al.* (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science*, **309**, 416–422.
9. Kabani,S., Fenn,K., Ivens,A., Ross,A., Smith,T.K., Ghazal,P. and Matthews,K.R. (2009) Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in *Trypanosoma brucei*. *BMC Genomics*, **10**, 427.
10. Siegel,T.N., Hekstra,D.R., Wang,X., Dewell,S. and Cross,G.A. (2010) Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic Acids Res.*, **38**, 4946–4957.
11. Jensen,B.C., Sivam,D., Kifer,C.T., Myler,P.J. and Parsons,M. (2009) Widespread variation in transcript abundance within and across developmental stages of *Trypanosoma brucei*. *BMC Genomics*, **10**, 482.
12. Queiroz,R., Benz,C., Fellenberg,K., Hoheisel,J.D. and Clayton,C. (2009) Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. *BMC Genomics*, **10**, 495.
13. Clayton,C.E. (2002) New EMBO member's review life without transcriptional control? From fly to man and back again. *EMBO J.*, **21**, 1881–1888.
14. Mahmood,R., Mittra,B., Hines,J.C. and Ray,D.S. (2001) Characterization of the Crithidia fasciculata mRNA cycling sequence binding proteins. *Mol. Cell. Biol.*, **21**, 4453–4459.
15. Pasion,S.G., Hines,J.C., Ou,X., Mahmood,R. and Ray,D.S. (1996) Sequences within the 5′ untranslated region regulate the levels of a kinetoplast DNA topoisomerase mRNA during the cell cycle. *Mol. Cell. Biol.*, **16**, 6724–6735.
16. Schurch,N., Furger,A., Kurath,U. and Roditi,I. (1997) Contributions of the procyclin 3′ untranslated region and coding region to the regulation of expression in bloodstream forms of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **89**, 109–121.
17. Vassella,E., Acosta-Serrano,A., Studer,E., Lee,S.H., Englund,P.T. and Roditi,I. (2001) Multiple procyclin isoforms are expressed differentially during the development of insect forms of *Trypanosoma brucei*. *J. Mol. Biol.*, **312**, 597–607.
18. Urwyler,S., Vassella,E., Van Den Abbeele,J., Renggli,C.K., Blundell,P., Barry,J.D. and Roditi,I. (2005) Expression of procyclin mRNAs during cyclical transmission of *Trypanosoma brucei*. *PLoS Pathog.*, **1**, e22.
19. Vassella,E., Den Abbeele,J.V., Butikofer,P., Renggli,C.K., Furger,A., Brun,R. and Roditi,I. (2000) A major surface glycoprotein of *Trypanosoma brucei* is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. *Genes Dev.*, **14**, 615–626.
20. Hehl,A., Vassella,E., Braun,R. and Roditi,I. (1994) A conserved stem-loop structure in the 3′ untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*. *Proc. Natl Acad. Sci. USA*, **91**, 370–374.
21. Furger,A., Schurch,N., Kurath,U. and Roditi,I. (1997) Elements in the 3′ untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation. *Mol. Cell. Biol.*, **17**, 4372–4380.
22. Walrad,P., Paterou,A., Acosta Serrano,A. and Matthews,K.R. (2009) Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA cis-elements. *PLoS Pathog.*, **5**, e1000317.
23. Mani,J., Guttinger,A., Schimanski,B., Heller,M., Acosta-Serrano,A., Pescher,P., Spath,G. and Roditi,I. (2011) Alba-domain proteins of *Trypanosoma brucei* are cytoplasmic RNA-binding proteins that interact with the translation machinery. *PLoS One*, **6**, e22463.
24. Hendriks,E.F., Robinson,D.R., Hinkins,M. and Matthews,K.R. (2001) A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form. *EMBO J.*, **20**, 6700–6711.
25. Hendriks,E.F. and Matthews,K.R. (2005) Disruption of the developmental programme of *Trypanosoma brucei* by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein. *Mol. Microbiol.*, **57**, 706–716.
26. Blackshear,P.J. (2002) Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. *Biochem. Soc. Trans.*, **30**, 945–952.
27. Carballo,E., Lai,W.S. and Blackshear,P.J. (1998) Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin. *Science*, **281**, 1001–1005.
28. Gao,G., Guo,X. and Goff,S.P. (2002) Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science*, **297**, 1703–1706.
29. Lai,W.S., Kennington,E.A. and Blackshear,P.J. (2002) Interactions of CCCH zinc finger proteins with mRNA: non-binding tristetraprolin mutants exert an inhibitory effect on degradation of AU-rich element-containing mRNAs. *J. Biol. Chem.*, **277**, 9606–9613.
30. Brun,R. and Schonenberger. (1979) Cultivation and in vitro cloning or procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. Short communication. *Acta Trop.*, **36**, 289–292.
31. Paterou,A., Walrad,P., Craddy,P., Fenn,K. and Matthews,K. (2006) Identification and stage-specific association with the translational apparatus of TbZFP3, a cch protein that promotes trypanosome life cycle development. *J. Biol. Chem.*, **281**, 39002–39013.
32. Aslett,M., Aurrecoechea,C., Berriman,M., Brestelli,J., Brunk,B.P., Carrington,M., Depledge,D.P., Fischer,S., Gajria,B., Gao,X. *et al.* (2010) TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.*, **38**, D457–D462.
33. Biebinger,S., Wirtz,L.E., Lorenz,P. and Clayton,C. (1997) Vectors for inducible expression of toxic gene products in bloodstream and procyclic *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **85**, 99–112.
34. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
35. Droll,D., Archer,S., Fenn,K., Delhi,P., Matthews,K. and Clayton,C. The trypanosome Pumilio-domain protein PUF7 associates with a nuclear cyclophilin and is involved in ribosomal RNA maturation. *FEBS Lett.*, **584**, 1156–1162.
36. Kramer,S., Kimblin,N.C. and Carrington,M. Genome-wide in silico screen for CCCH-type zinc finger proteins of *Trypanosoma brucei, Trypanosoma cruzi and Leishmania major*. *BMC Genomics*, **11**, 283.
37. Bastin,P., Bagherzadeh,Z., Matthews,K.R. and Gull,K. (1996) A novel epitope tag system to study protein targeting and organelle biogenesis in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **77**, 235–239.
38. Kramer,S., Queiroz,R., Ellis,L., Webb,H., Hoheisel,J.D., Clayton,C. and Carrington,M. (2008) Heat shock causes a decrease in polysomes and the appearance of stress granules in trypanosomes independently of eIF2{alpha} phosphorylation at Thr169. *J. Cell Sci.*, **121**, 3002–3014.
39. Cassola,A., De Gaudenzi,J.G. and Frasch,A.C. (2007) Recruitment of mRNAs to cytoplasmic ribonucleoprotein granules in trypanosomes. *Mol. Microbiol.*, **65**, 655–670.
40. Beissbarth,T. and Speed,T.P. (2004) GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics*, **20**, 1464–1465.
41. De Gaudenzi,J., Frasch,A.C. and Clayton,C. (2005) RNA-binding domain proteins in Kinetoplastids: a comparative analysis. *Eukaryot. Cell.*, **4**, 2106–2114.
42. Archer,S.K., Luu,V.D., de Queiroz,R.A., Brems,S. and Clayton,C. (2009) *Trypanosoma brucei* PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. *PLoS Pathog.*, **5**, e1000565.
43. Estevez,A.M. (2008) The RNA-binding protein *Tb*DRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes. *Nucleic Acids Res.*, **36**, 4573–4586.
44. Dean,S.D., Marchetti,R., Kirk,K. and Matthews,K. (2009) A surface transporter family conveys the trypanosome differentiation signal. *Nature*, **459**, 213–217.

45. Szoor,B., Wilson,J., McElhinney,H., Tabernero,L. and Matthews,K.R. (2006) Protein tyrosine phosphatase *Tb*PTP1: a molecular switch controlling life cycle differentiation in trypanosomes. *J. Cell Biol.*, **175**, 293–303.
46. Alsford,S., Turner,D.J., Obado,S.O., Sanchez-Flores,A., Glover,L., Berriman,M., Hertz-Fowler,C. and Horn,D. High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res*, **21**, 915–924.
47. Lahav,T., Sivam,D., Volpin,H., Ronen,M., Tsigankov,P., Green,A., Holland,N., Kuzyk,M., Borchers,C., Zilberstein,D. *et al.* (2011) Multiple levels of gene regulation mediate differentiation of the intracellular pathogen *Leishmania*. *FASEB J.*, **25**, 515–525.
48. Rosenzweig,D., Smith,D., Opperdoes,F., Stern,S., Olafson,R.W. and Zilberstein,D. (2008) Retooling *Leishmania* metabolism: from sand fly gut to human macrophage. *FASEB J.*, **22**, 590–602.
49. Ling,A.S., Trotter,J.R. and Hendriks,E.F. (2011) A zinc finger protein, *Tb*ZC3H20, stabilizes two developmentally regulated mRNAs in trypanosomes. *J. Biol. Chem.*, **286**, 20152–20162.